

Cotton, Christopher; Nordstrom, Ardyn; Nanowski, Jordan; Richert, Eric

Working Paper

Can dialogues around girls' education improve academic outcomes? Evidence from a randomized development project

Queen's Economics Department Working Paper, No. 1492

Provided in Cooperation with:

Queen's University, Department of Economics (QED)

Suggested Citation: Cotton, Christopher; Nordstrom, Ardyn; Nanowski, Jordan; Richert, Eric (2022) : Can dialogues around girls' education improve academic outcomes? Evidence from a randomized development project, Queen's Economics Department Working Paper, No. 1492, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:

<https://hdl.handle.net/10419/281096>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Can Dialogues Around Girls' Education Improve Academic Outcomes? Evidence from a Randomized Development Project

Christopher S. Cotton, Ardyn Nordstrom, Jordan Nanowski, and Eric Richert

ABSTRACT. We evaluate the impact of deliberative dialogues about girls' education with groups of parents, teachers, and girls on the education outcomes of girls in rural Zimbabwe. Dialog-based engagement campaigns increased mathematics performance and school enrolment. In later periods, the program was expanded to provide resources and an updated curriculum. During these later periods, we observed improvements in literacy, but no additional improvements in mathematics and enrolment beyond what was observed following the dialog-based engagement campaign alone. A mediation analysis shows how earlier gains in math performance due to the dialogues are positively associated with later gains in literacy.

Date: **November 2022.**

Key words and phrases. Girls' Education Challenge, education, impact evaluation, randomized controlled trial, multifaceted intervention, community conversations, deliberate dialogues, social norms

JEL: I24, I25, J16, O12, O15.

Cotton (cc159@queensu.ca): Queen's University and Limestone Analytics. **Nordstrom** (ardyn.nordstrom@carleton.ca): Carleton University and Limestone Analytics. **Nanowski:** Canadian Mortgage and Housing Corporation. **Richert:** Princeton University and University of Chicago.

Acknowledgements: The data collection was funded by the UK's Foreign, Commonwealth & Development Office (FCDO/UKAid) and conducted by Miske Witt & Associates in collaboration with World Vision and PricewaterhouseCoopers LLC, the UKAid Girls' Education Challenge fund manager. The dataset was provided by World Vision to Cotton for use in academic research and is available to others by request through World Vision UK or FCDO/UKAid. A preliminary exploration of the data by the research team was funded by World Vision Canada and conducted through Limestone Analytics, where Cotton and Nordstrom hold secondary affiliations and Nanowski previously worked. The project received General Research Ethics Board approval at Queen's University (project GECO-008-19) for the secondary use of data. The research team greatly appreciates discussions with Bahman Kashi, Terry Grey, Mbuso Jama, Janelle Zwier, and the IGATE project team at World Vision. Nordstrom and Richert are grateful for graduate studies funding through the Ontario Graduate Scholarship and the Social Sciences and Humanities Research Council. Cotton is grateful for financial support provided through his position as the Jarislowsky-Deutsch Chair in Economic & Financial Policy at Queen's University. The paper has benefited from comments from seminar participants at several universities.

1. INTRODUCTION

Youth around the world face significant challenges in their pursuit of education, with girls, on average, facing greater barriers than boys. Girls often receive less schooling and perform worse in certain fields, such as mathematics, compared to boys.¹ The relative performance of girls tends to be especially low in settings where there are persistent beliefs or social norms that undervalue girls' education.² Stereotypes, including beliefs around mathematics ability, not only exist among parents and teachers (e.g. Robinson-Cimpian et al., 2014; Alan et al., 2018; Lavy & Sand, 2018), but also among students themselves, potentially contributing to differences in effort and interests (e.g. Dhar et al., 2022; Bian et al., 2017).³ Furthermore, in many settings, parents, teachers, youth, and others often do not understand the benefits of education, especially for girls (Jensen, 2010; Attanasio & Kaufmann, 2014).

Improving education outcomes for girls is particularly challenging in communities where parents, teachers, or students undervalue girls' education, or where negative stereotypes, biases, or social norms present barriers to girls' education. Recognizing this, when designing gendered education projects, development agencies such as the U.K.'s Foreign, Commonwealth & Development Office (FCDO/UKaid) and the U.S. Agency for International Development (USAID) frequently incorporate interventions that use discussions to help shape the way that parents, teachers, and students think about the barriers to and benefits from girls' education and girls' rights (GEC, 2018; Cislighi et al., 2019). Such broadly defined "*dialog-based engagement*" campaigns combine information provision and group discourse in an attempt to build recognition of the issues and develop willingness to address them by members of the communities. Such forms of

¹For a detailed description of recent gender gaps across countries, see OECD (2015). A gender gap in mathematics performance exists across many countries (Guiso et al., 2008) and within every stratum of society (Fryer & Levitt, 2010). For a discussion of how gender gaps in schooling have been reduced in some environments including North American tertiary education, see Goldin et al. (2006); Hyde et al. (2008); Asadullah & Chaudhury (2009), and Rosenzweig & Zhang (2013). Many research efforts have worked to understand the barriers to girls' education that contribute to these and other performance gaps (e.g., Benbow & Stanley, 1980; Hedges & Nowell, 1995; Duflo, 2012).

²Girls perform relatively worse in countries with less gender equality (Guiso et al., 2008; Chang et al., 2020), and in North America gaps are largest among students whose parents come from less-equitable cultures or countries with less female empowerment (Nollenberger et al., 2016; Rodriguez-Planas & Nollenberger, 2018). It has also been shown that the gender stereotypes and biases of teachers are associated both with lower test performance of girls and minorities (Robinson-Cimpian et al., 2014; Alan et al., 2018), and with long-run schooling attainment and career choices (Lavy & Sand, 2018). See also Burgess & Greaves (2013) for related analysis involving racial minorities.

³In summarizing the recent data and literature, the OECD (2015) argued that "gender disparities in performance do not stem from innate differences in aptitude, but rather from students' attitudes towards learning and their behavior in school, from how they choose to spend their leisure time, and from the confidence they have - or do not have - in their own abilities as students..." and that improving outcomes demands the greater involvement of parents, teachers, and students themselves.

deliberative dialogue are ubiquitous in the projects of many development organizations, but are rarely implemented in isolation, with organizations adopting them as one component of many multi-component projects. In education projects, they accompany incentives, financing, learning resources, teacher training, or curriculum changes in support of education.

When a dialogue-based engagement intervention is implemented as a single piece of a larger program, it is typically not possible to identify to which extent it contributed to the program's overall impact, and there remains little quantitative evidence of their effectiveness (GEC, 2018; Cislighi et al., 2019). The Girls' Education Challenge (GEC), a £500 million UKaid development challenge fund, has financed the implementation of 41 major gendered education projects across 17 developing countries. In 2018, despite prioritizing data collection and evaluation and despite including some version of dialogue-based engagement as a component in most of its evaluations, the GEC concluded that there remains little evidence about the impact of such engagement efforts. There is still "a need for more detailed insights into how targeting community attitudes and behaviors on their own can affect learning, attendance and retention of girls" GEC (2018).

The current paper addresses this evidence gap using data from one of the GEC's multifaceted education programs implemented from 2014-2016 in rural Zimbabwe. We provide the first robust, quantitative evidence that dialogue based engagement campaigns in any setting can have meaningful impacts on behavior and outcomes. Unlike other GEC programs, the "Improving Girls' Access through Transformative Education" (IGATE) project implemented dialogue-based engagement components in a staggered roll-out that allows us to not only assess the overall impact of the IGATE program but also to identify the impact of the dialogue-based engagement campaigns, highlighting how much these components contributed to the multifaceted program's overall impact.

The dialog-based engagement campaigns used by IGATE and other development projects go beyond providing people with facts like the expected financial returns from additional schooling (e.g. Nguyen, 2008; Jensen, 2010) or the benefits of chlorinating water (e.g. John & Orkin, 2022).⁴ Rather, dialog-based approaches encourage participants

⁴Other studies of information provision have considered the impact on academic performance from providing parents information about student performance (Berlinski et al., 2016; Dizon-Ross, 2019; Barrera-Osorio et al., 2020; Doss et al., 2018), or school quality (Andrabi et al., 2017; Banerjee et al., 2010; Hastings & Weinstein, 2008). It also goes beyond combining facts and suggestions on how to apply the information or what actions can be taken to address issues, such as in Cortes et al. (2018); Doss et al. (2018). Bettinger et al. (2012) explores the role of application assistance and information provision to parents on college applications. Additionally, (Oreopoulos et al., 2017; Lavecchia et al., in press) and (Walsh et al., 2014) examine youth mentorship programs that share some similarities to IGATE. Our analysis is also related to studies considering how information provision impacts other non-education outcomes in developing countries. Information-based interventions have been particularly successful in improving health outcomes. Such

to actively discuss the issues with the facilitators and with the other participants. With IGATE, the discussions focused on the benefits of girls' education, and the specific barriers to girls' education in that community and involved groups of mothers, fathers, girls, teachers, and other members of the community. For example, during a mothers group meeting, participants may focus on raising awareness of factors affecting girls' education and well-being by discussing topics such as girls' responsibilities at home, and how gender-based factors influence how chores are distributed within households. The groups also encouraged participants to discuss the topics on their own with others in their community. This kind of encouraged diffusion is a common feature of dialog-based interventions (Cislaghi et al., 2019), and was encouraged by IGATE facilitators.⁵

The dialog-based engagement campaigns can be a more effective way of communicating information, allowing facilitators and other participants to discuss concerns, personal experiences, and potential solutions, and allowing participants to observe support for the issues among their peers. This can encourage open discussion of the topics within communities, change beliefs, and create momentum for changing behavior and social norms.⁶ Furthermore, while the dialog-based campaign was intended to create an empowering and engaging environment for girls education, this dialog-based approach is different from the more common empowerment training offered by the Bangladesh Rural Advancement Committee (BRAC) and others, which typically emphasize empowerment through skills-based training (e.g. Bandiera et al., 2020; Acevedo et al., 2017).

After the initial dialog-based engagement campaigns were established, the IGATE program then expanded the programming to introduce support for teachers in schools and to provide books to classrooms and bicycles to girls living far from school. It is important to note that the dialog-based engagement campaign was not designed to collect information affecting the design of these later programs. Rather, the additional supports were originally intended to begin along side the dialog-based engagement campaign but were stalled due to administrative delays. The unique feature of IGATE that facilitates

issues have been considered in the context of safe-sex practices (Dupas, 2011), and breastfeeding and nutrition (Fitzsimons et al., 2016; Krämer et al., 2019), for example. Additionally, information interventions have been shown to increase the number of small businesses that receive a loan (De Mel et al., 2011) and to increase labor mobility for workers in poor work environments (Shrestha & Yang, 2019).

⁵There is strong qualitative evidence that such diffusion was successfully taking place in the IGATE program. For example, one community leader remarked that "As women attend those [IGATE] meetings they pass the same information to those at home and that's how the information is cascading." when asked if the discussions from IGATE were being shared with the community.

⁶See, for example, Cook et al. (2017), which examines the impact of community dialogues that focused on race and ethnicity within the education community. The potential to influence social norms is evident in the IGATE project's qualitative data, where one parent specifically noted that "as members of the mothers group, we have also taken advantage of community meetings to educate [others in] our community; particularly parents on the benefits of educating the girl child. I would like to believe that this has accounted for the changing attitudes towards girl child's education by parents in our community."

our analysis is that it collected data in treatment and control communities (it's "midline" evaluation) after the program widely implemented the dialogue-based engagement campaigns, but before it implemented the other program components. This was in addition to baseline and endline data collection that took place before and after the project was complete. The staggered implementation, and its alignment with data collection points, allows us to identify the impact of the dialogue-based engagement efforts on the academic performance of girls, before other components of the program were implemented.

Our analysis shows that these dialog-based engagement campaigns resulted in an improvement in mathematics performance and school enrolment within a relatively short time frame. The improvement in mathematics performance occurred even though the dialog-based engagement campaign did not specifically focus on math or STEM participation or performance. There are several reasons that a general campaign to encourage girls' education may have such an impact on math performance. For example, the campaign may have encouraged greater effort or focus by girls on tasks that others have shown are sometimes not prioritized because they are often viewed as difficult, masculine, and largely irrelevant (Gudyanga, 2016). It may also have led to increased teacher or parental attention for girls after community members gained awareness of gender gaps. We see no similar improvement in literacy in the months after girls and communities were exposed to only the dialog-based engagement campaign.

After isolating the impact of the dialog-based engagement campaign, we then examine the ultimate impact of the entire IGATE project. Overall, the combined set of IGATE interventions led to significant improvements in literacy among the girls in the treatment communities. The later intervention components of the project, including the provision of books and teacher support, likely contributed to the observed improvements in literacy. A formal mediation analysis suggests that the later gains were at least partially associated with the program's earlier impact achieved by midline. Our results suggest that the dialog-based engagement campaign was the main driving force behind the project's early impact on math performance and enrolment, but that these conversations alone was not sufficient to improve literacy outcomes in the time period considered.

Compared to past research on the impact of engagement or information-based interventions, our study is novel on several dimensions. This is the first study to isolate the impact of a dialog-based engagement campaign from a major education-focused development aid project. This provides insight into a type of information provision that is very popular in practice, but difficult to study as it is rarely implemented in isolation. The type of information provided and then discussed in IGATE emphasizes the rights of marginalized girls to pursue education, and highlights the general importance of additional schooling at the primary level. These discussions took place separately with

groups of girls, parents, teachers, and other community members to raise awareness of both the benefits of education and the types of barriers girls face in pursuit of education. We show how such engagement has successfully changed education outcomes for girls, resulting in relatively substantial gains in mathematics performance relatively quickly, and contributing to longer term improvements in literacy.

The paper proceeds as follows. Section 2 describes the IGATE project. Section 3 describes the randomization of treatment, the process through which data was collected and literacy and numeracy performance measured, and the estimation strategy. Section 4 presents the results. Section 5 concludes.

2. CONTEXT AND PROGRAM DESCRIPTION

In 2012, UKaid⁷ launched its GEC initiative, a multi-year commitment to improving access to education and learning for marginalized and at-risk youth, especially girls, in the developing world. The GEC is the largest-ever donor-funded program focused on girls in developing countries. The GEC has spent roughly £500 million in support of dozens of projects across the developing world and claims to have directly benefited more than a million girls.⁸

The GEC's IGATE project was implemented between 2014 and 2016 by a consortium of nongovernmental organizations (NGOs) led by World Vision.⁹ The program was implemented in randomly-selected schools across 10 primarily-rural districts in Zimbabwe and is estimated to have reached a total of 48,773 girls.

The project adopted a multifaceted intervention design in an effort to increase the accessibility and quality of education for marginalized girls. The initial wave of the project consisted of the dialog-based engagement campaign, which involved voluntary meetings with groups of girls, parents, teachers, and other community members to discuss issues related to girls' education. During these meetings, community members were guided through discussions on girls' rights, the importance of girls attending school, and the barriers that girls and their families face in pursuing education. Participants discussed how they could specifically support girls' education by, for example, supporting the enrolment and attendance of girls in school and increasing awareness around specific barriers faced by girls including gender-based violence, unequal expectations

⁷UKaid is funded by the U.K.'s Foreign, Commonwealth, and Development Office (FCDO). At the time of implementation, the U.K. donor agency was referred to as the Department for International Development (DFID).

⁸More details on the Girls Education Challenge can be found at <https://girlseducationchallenge.org/>.

⁹Partner organizations included CARE International, SNV Netherlands Development Organisation, Emthonjeni Women's Forum, Happy Readers, World Bicycle Relief, and the Union for the Development of the Apostolic Church in Zimbabwe Africa.

around household or classroom chores, menstrual hygiene needs, and limited household resources. While these community members could have met on their own, or in other venues the organized sessions provided a setting for participants to comfortably discuss a guided set of topics in the presence of a trained facilitator. These sessions provided suggestions, but no direct support, for how communities, schools, and families could mitigate some of these barriers.

Later, after midline data collection, the program was expanded to provide teacher support, learning resources, and books in the local primary schools, and to provide bicycles to girls who lived far from school.

The IGATE dialog-based engagement campaigns included a broad set of discussion groups focusing on different members of the community. Unfortunately we cannot separately identify the effects of these individual treatments. This means we cannot examine if engaging girls or parents were more important for the overall effect of the program. This is because all five groups were meeting regularly in all the treated locations before the midline follow up. The dialog-based campaign included the following specific intervention components:

- **Community in Support of Girls' Education (CSGE):** Implemented by Government of Zimbabwe employees in the Ministry of Primary and Secondary Education trained by IGATE staff, CSGE promoted girls' education throughout the communities by providing participants an opportunity to learn about and discuss girls' education and the barriers they face in their pursuit of education. It also provided participants with information about the minimum standards that could be expected from local primary and secondary schools. During these meetings, participants discussed how their communities could hold schools accountable for providing quality education for all students. An average of 201 individuals participated in CSGE meetings in each treatment community.
- **Mothers Groups (MGs):** The project recruited local female caregivers to participate in MGs and led discussions within these MGs about the importance of girls' education and school attendance. Participants were encouraged to share this information throughout their community. These groups also highlighted the challenges girls face due to gender-based violence, inequitable treatment, and hygiene and menstruation. The mothers were then provided guidance on how to mentor girls on these topics and training on how to make reusable menstrual pads. In some places, fathers also participated in these groups. An average of 15 mothers and 5 fathers participated in MGs in each treatment community.

- **Power Within Clubs (PWCs):** The project recruited teachers to set up and run PWCs within schools. These teachers were mentors for each club and were often linked with the local MG. The PWCs were designed to encourage girls' to take an active role in decisions around their education by discussing girls' rights and how to potentially navigate barriers to girls' education. During PWC meetings, participants were guided through discussions on the importance of education, attending school, and doing school work. As in the MGs, participants were encouraged to share their knowledge with others in the school who did not directly participate in the groups. An average of 41 girls participated in PWCs in each treatment school.
- **Village Savings and Loan (VSL):** The project led savings groups operating in the communities through discussions that focused on the importance of encouraging girls' education while providing suggestions for how participants could save for future education needs. The project also provided guidelines for community members to establish such groups in accordance with CARE International's VSL model. It did not provide any financial assistance or other resources to these groups or their members. An average of 49 adult community members participated in VSL groups in each treatment community.
- **School Development Committees (SDCs)**—These school-based committees provide teachers and school officials with information and opportunities to discuss the importance of creating learning environments that were gender sensitive. This included providing guidance on how schools and teachers could support MG efforts relating to hygiene and menstruation. An average of 8 individuals participated in SDCs in each treatment school.

Following the community-wide information and dialogue-based engagement efforts, the IGATE project expanded to include other interventions that were not focused on information provision. This second stage of the project involved providing bicycles to girls with long commutes to school,¹⁰ and a *Happy Readers* program, based in part on World Vision's widely-implemented Unlock Literacy program, which provided literacy and reading materials to schools to help students learn to read.¹¹ When the program expanded after midline to provide books and bikes, an average of 96 bikes and 1,478

¹⁰These bicycles were provided through the Bicycle Education Empowerment Program (BEEP) in partnership with the World Bicycle Relief organization.

¹¹There were a small number of locations where girls received bicycles before midline data was collected. Discussions with the project suggest that there was no systematic reason that some schools received bicycles early and that this is not associated with either the need for bicycles or the expected impact of bicycles. These locations have therefore been dropped from this analysis to isolate the impact of the information-based interventions.

books were delivered at each treatment school. These resource-based interventions were originally intended to be rolled out at the start of the program but were stalled due to administrative delays in approvals for budget transfers.

As with the GEC initiative more generally, IGATE took a multifaceted approach to improve girls' education outcomes. Its project design builds on a theory of change in which sustainable impact requires changing attitudes, beliefs, and social norms within communities, and that the greatest impact will be had by projects that work to change attitudes while also working to increase resources, or improve infrastructure, teaching practices, policy, or institutions. See (Unterhalter et al., 2014) for the UKaid review of the suggestive evidence in support of such an approach undertaken at the beginning of the GEC. The atypical feature of the IGATE project compared to other GEC projects was not in its inclusion of dialogue-based efforts to change attitudes, but in the way that the timing of implementation allows us to isolate the impact of these intervention components.

2.1. Dates. The data collection and program timing are summarized in Figure 1. Baseline data collection occurred before implementation began in February 2014. Midline data collection took place a year and a half later in June-August 2015. At this point, the community dialog-based engagement campaign had been widely implemented, but the non-information interventions had not. Before midline data was collected, each of the treatment locations in the sample had received all five of the community information-based interventions. After midline data collection the project continued the information and engagement campaign and introduced non-information interventions. Endline data collection occurred in November-December 2016 at the end of the project. All collection dates occur within the middle of the school semesters.

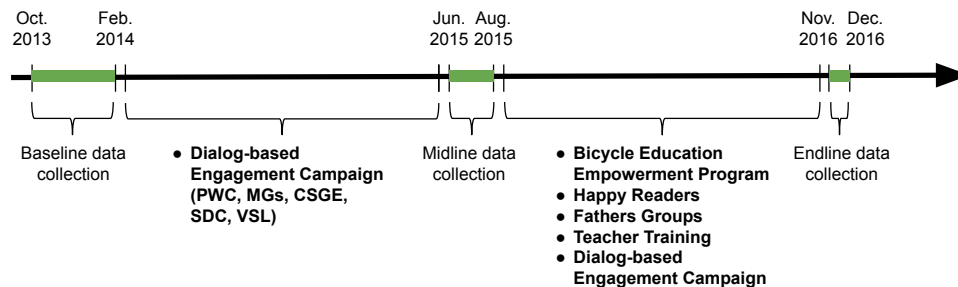


FIGURE 1. IGATE Data Collection and Program Implementation Timeline

2.2. Dialogue-driven change. The dialogue-based engagement campaign implemented by the IGATE program builds on implementing partner CARE International's widely-used Social Analysis and Action (SAA) methodology for encouraging shifts in attitudes

and actions. The SAA methodology (CARE, 2014) explains that “the goal of facilitation is not to proselytize or promote certain attitudes, norms, and behaviors, but rather to facilitate a process of dialogue and reflection that ... aims to surface individual and community attitudes and norms, explore how these attitudes, norms and power dynamics shape individual and collective behavior, and reflect on the consequences of those norms and behaviors.” It formulates three core elements, involving leading groups of participants through (1) reflection, dialogue, and exploration of ideas, (2) envisioning alternatives and challenging harmful norms, and (3) commitment to action for improved outcomes.¹²

For example, the group might be asked to reflect on the reasons girls in their communities may not attend school. Based on this discussions, facilitators would encourage the participants to reflect on what contributes to these factors being barriers. For example, if the group identifies that early marriage as a risk factor affecting girls' attendance, the groups would then reflect on the factors that contribute to this. In response to these reflections, the group would discuss ways to improve the girls' environment by, for example, partnering with religious leaders to raise awareness of the harms associated with early marriage practices.

The SAA model is closely related to other theories of change through which dialogues are expected to altering social norms and related behavior. In early work, Chasin et al. (1996) argued that public dialogues structured around family therapy sessions could help groups with potentially diverging opinions talk about and reduce social conflict on divisive issues. While the initial efforts in this space largely focused on reducing conflict around race and ethnicity,¹³ the dialogue-based engagement process is now frequently used to build consensus for change and collective action within communities. For example, Schweizer et al. (2013) discuss dialogues as a theoretical framework for engaging communities on climate change issues; and Martin et al. (2017) consider how dialogues may change caregiver practices in support of child health and nutrition. Although much of the research in this space draws favorable conclusions about the ability of dialogue to change outcomes, little of this research is causal or quantitative in nature. Most of the research in this area presents theoretical arguments (e.g., Schweizer et al., 2013), qualitative assessments (e.g. Martin et al., 2017), or before-and-after analyses without control groups (e.g., Ndagije et al., 2019; Wegs et al., 2019). Figueroa et al. (2016) studies whether community dialogues change cultural norms for HIV prevention in Mozambique by comparing reported beliefs of a recent program cohort after participation and

¹²See also CARE's SAA implementation guide, which walks through the theory of change in detail (CARE, 2007).

¹³See Dessel et al. (2006) for examples of such efforts implemented in communities.

an upcoming program cohort before participation, showing beneficial effects on gender beliefs, gender roles, and HIV stigma.

Our analysis complements the existing literature by providing a causal analysis of the impact of dialogue-based engagement on girls' education outcomes. It is based on the evaluation of treatment and control data from a large-scale randomized intervention with a staggered roll-out of components that allows us to estimate the impact of a dialogue campaign on education outcomes, presenting some of the most robust evidence as to the meaningful impacts that such efforts can have on outcomes.

3. DATA AND METHODOLOGY

3.1. Sampling Framework. The program identified 467 schools (and respective communities) across rural districts in Zimbabwe to be eligible to receive the IGATE interventions. A subset of these schools was randomly selected to receive the program.¹⁴ As is standard in GEC evaluations, only a subset of the participating schools were then selected for data collection and evaluation, providing a sample of 37 treatment and 28 control locations from the eligible primary schools. This trial was registered with the American Economic Association's registry for randomized controlled trials.^{15 16}

Data collection was conducted by teams of professional enumerators from a Zimbabwe-based firm. All primary school-aged girls in the local communities were eligible for inclusion in our sample, regardless of whether or not they or any family member directly participated in the IGATE program.¹⁷ That is, the sample defines the treatment status of girls based on whether their location received IGATE rather than whether they were direct participants. To select girls and their caregivers for the data collection, enumerators would start at a recognizable local landmark in IGATE communities (e.g., community center) and then walk in different directions and knock on every fourth door. At the

¹⁴The map in Figure A1 in the appendix shows the location of treatment and control schools across Zimbabwe.

¹⁵The registry record can be found here: <https://www.socialscisearch.org/trials/7963>.

¹⁶The evaluation focused on the schools that were not part of the conflicting GEC Campaign for Female Education ("CamFed") initiative, which focused on secondary education. The GEC did collect additional data from the CamFed-eligible schools. However, this will not be included in this analysis since it is beyond the scope of the current research, which is focused on the impact of IGATE interventions alone. There is no association between IGATE treatment status and the rollout of the CamFed program, so this is not a concern for our identification strategy.

¹⁷GEC data collection initially included some secondary-school-aged girls in the sample. For clarity of interpretation, we limit our sample to girls who were in grade seven or below at baseline (i.e. in primary school at baseline). We do this for several reasons. First, IGATE was a primary-school-focused program so we wouldn't expect the same impact on secondary-aged students as on primary-aged students. Second, it was not possible to accurately determine the extent that secondary school students received access to program components. And third, the small sample size for secondary school students prevented subgroup analyses, meaning that their inclusion would likely bias the estimates in unknown ways.

household, they then would ask if any girls from IGATE schools of the appropriate age lived there. If an eligible girl lived there the enumerators proceeded to do the questionnaires, returning later if the girl was not home.

There are 1,160 girls observed in our sample at baseline, 942 at midline, and 820 at endline. For each school in the panel data set, there are between three and twenty girls, with an average of 11 girls from each location. For each girl in the data set, we have information provided by their caregivers and teachers, as well as data from standardized reading and mathematics tests. The attrition rate was similar across treatment and control locations with rates of 21% and 16%, respectively, over the three years between baseline and endline data collection.¹⁸ Although this is relatively large, it is consistent with other studies in similar contexts and is likely driven by the degree of migration present in the general population.

In the appendix, we compare the characteristics and test scores of the sample by attrition during the study (see tables A1 and A2 in the Appendix). Given the ex-ante similarity of the girls who dropped out of the sample, we are not concerned that girls that attrited from the sample are systematically different from girls who remain at midline or endline in any way that will clearly bias the analysis. Across the main household controls and test scores, the mean scores and standard deviations in the two groups are similar and well within a standard deviation of each other, which provides summary statistics of baseline test scores.

At midline, there are 385 and 557 girls in the control and treatment samples, respectively, which we define as our main sample. The treatment and control groups in this main sample exhibit similar observable characteristics at baseline, as shown in Table 1. The groups also have similar baseline test scores and grade distributions (as shown in Tables A3 and A4 in the Appendix).

3.2. Learning Assessments. The main data collected at baseline, midline, and endline included a survey with girls, a survey with a girl's caregiver, and the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA). Originally designed for the United States Agency for International Development (USAID), EGRA has been used to assess reading skills in primary school-aged students in over 70 countries and by hundreds of projects worldwide. EGMA, which was developed after EGRA, has been used to assess mathematics skills in primary school-aged students in over 20 countries around the world. In one study by Friedman et al. (2016), EGRA and EGMA

¹⁸Note that girls who had dropped out of school were still included in the sample, provided they could be located. Girls who have dropped out of the sample should not be assumed to be out of school since they may have moved to a new school in a different region.

TABLE 1. Baseline Summary Statistics

	Control	Treatment	Difference
Age	9.377 (2.011)	9.380 (1.997)	0.003
Grade	3.676 (1.746)	3.627 (1.779)	-0.049
Illness	0.103 (0.305)	0.112 (0.316)	0.009
Disability	0.174 (0.380)	0.191 (0.394)	0.017
Orphan	0.0676 (0.252)	0.0536 (0.226)	-0.014
Travel time to school (minutes)	32.96 (23.16)	35.12 (27.30)	2.16
Household often goes hungry	0.224 (0.418)	0.228 (0.420)	0.004
Household often goes thirsty	0.128 (0.335)	0.121 (0.327)	-0.007
Caregiver has no education	0.0676 (0.252)	0.0979 (0.298)	0.0303
Caregiver has primary education	0.523 (0.500)	0.550 (0.498)	0.027
Caregiver has secondary education	0.409 (0.493)	0.352 (0.478)	-0.057
Caregiver works outside of household	0.221 (0.415)	0.235 (0.425)	0.014
N	385	557	

Note these numbers measure baseline levels for girls who could be recontacted at midline.

were found to be the most frequently used assessment systems in education evaluations in Eastern and Southern Africa.

The version of EGMA implemented here included the common subtasks of number identification, quantity discrimination, missing numbers, basic addition, and basic subtraction. The version of EGRA implemented here involved five subtasks: letter sound identification, invented word reading, reading fluency, and reading comprehension. We provide a detailed description and examples of EGMA and EGRA subtasks in the appendix.

At midline, similar tests were administered with slight variations from the baseline versions. It is necessary to change the versions of the tests to separate learning from recall in the analysis. However, the EGRA and EGMA subtasks have strict development guidelines that ensure the difficulty level is standardized across versions.

3.3. Econometric Strategy. We test our hypothesis that the IGATE treatment led to improvements in learning, enrolment, and caregiver behavior by estimating the following ordinary least squares regression at both midline ($t = 1$) and endline ($t = 2$):

$$Y_{ijt} = \beta_0 + \beta_1 Treated_{ij} + \beta_2 X_{ij0} + \beta_3 Y_{ij0} + \epsilon_{ijt} \quad (1)$$

Where Y_{ijt} represents the outcome variable (learning, enrolment, behavior indicators) measured at either midline or endline, respectively. By extension, Y_{ij0} represents the outcome variable measured at baseline.¹⁹ $Treated_{ij}$ represents whether girl i is in a school j that was exposed to IGATE. And X_{ij0} includes a set of controls for the girl's age, grade, and indicators for the district $girl_i$ lives in. Finally, ϵ_{ijt} is the disturbance term. In all of our estimations, standard errors are clustered at the school level. Since our treatment was applied at the school level, and participation in the interventions was voluntary, this specification provides estimates of the project's intent-to-treat effect.

Given the importance of understanding the sequential impacts of the interventions, we included estimations of a modified version of equation 1, which includes midline test scores when we are assessing endline test score outcomes:

$$Y_{ij2} = \alpha_0 + \alpha_1 Treated_{ij} + \alpha_2 X_{ij0} + \alpha_3 Y_{ij0} + \alpha_4 Y_{ij1} + \epsilon_{ij2} \quad (2)$$

When combined with the results for 1, this allows us to conduct a mediation analysis following the approach described in Acharya et al. (2016), as well as Judd & Kenny (1981); James & Brett (1984); Baron & Kenny (1986); VanderWeele (2016).²⁰ By controlling for midline test scores we are able to determine if the impact on test scores at midline mediates the ultimate impact observed at endline.

The randomization of treatment locations allows for an experimental approach to the evaluation. External factors such as the presence of droughts or other policy changes that occurred are likely to affect both the treatment and control groups similarly. In this way, the evaluation can be considered a causal analysis, identifying the changes in outcomes attributable to the IGATE project. Since participation in the treatments was voluntary, we focus on estimating intent-to-treat effects for our analysis.

Since there were administrative issues that prevented the non-information interventions to be implemented until after midline, this provides a unique opportunity to isolate the impact of the dialog-based engagement campaign at midline. While it is plausible that the expectation of additional future benefits may have increased participants' willingness to participate in the information-based interventions, these future possible

¹⁹By design, all girls are enrolled in school at the beginning program. Therefore, this term is not included in the specification where Y_{ijt} is enrolment. However, this lagged-dependent value specification is particularly important for test scores, as shown in Singh (2020).

²⁰More details on the mediation analysis are provided in Appendix C

benefits were not promoted widely before midline, so we do not expect this anticipation to significantly affect the results.

In practice, the dialog-based intervention is often combined with other programs. This has limited the ability of researchers to isolate the impact of these interventions. While the variation in timing of the roll-out in the IGATE project allows us to identify the effect of this program when implemented alone the effectiveness of the dialog based campaign could interact with those from the other programs. It seems likely that these effects would be complementary, for example when girls face lower barriers and are better supported in their education improved classroom resources could be made more effective and so our estimates may underestimate the impact of the dialog-based programs when implemented in conjunction with a broader set of interventions.

The impact of IGATE measured at endline cannot be attributed to any subset of the program interventions as all components were implemented before that time. The initial information campaigns may have contributed to greater awareness among community members to ultimately increase the productivity of the second set of interventions introduced after midline. Indeed, the mediation analysis presented in the results that the improvements observed in numeracy at midline at least partially, suggesting that the project's earlier impact allowed students to make further learning gains.

The analysis focuses on the impact of the IGATE program on numeracy and literacy (as measured by EGMA and EGRA), as well as enrolment. To provide additional insights, we also consider the impact of the project on specific subtasks of the learning assessments, and the number of questions attempted.

4. RESULTS

The IGATE program's primary aim is to improve access to quality education for marginalized girls. Our evaluation assesses the impact of the IGATE program on enrolment, and performance in mathematics and literacy.²¹ We also consider the impact on the frequency of caregivers' supportive actions towards girls.

4.1. Enrolment. To look at the accessibility of education after IGATE, Table 2 shows that girls in treatment areas are 2.5 percentage points more likely to be enrolled in school than girls who did not receive treatment by midline. This difference between baseline and midline is statistically significant and also intrinsically meaningful when dealing with a primary-school-aged population for which enrolment rates are high. Specifically, at baseline over 99.0% of girls in the sample across treatment and control areas were

²¹As shown in Nordstrom & Cotton (2020), the impact on enrolment and learning can move in opposite directions in response to external factors, potentially leading to misleading conclusions about the benefits of a program to education outcomes. This motivates the evaluation of both enrolment and learning outcomes to confirm no adverse consequences to education overall.

enrolled in school. At midline, this fell to 96.1% in the control group, but only fell to 98.5% in the treatment group. This represents a substantial decrease in the dropout rate, with the dropout rate in the control group being more than five times that in the treatment group between baseline and midline. This is evidence that the IGATE program led to substantial changes in the education attainment of those who would have left school in the absence of the program, who are arguably the most marginalized of the girls in the sample.

Because IGATE provided only the information-based interventions to the treatment communities ahead of midline, the entirety of the program's impact on enrolment at midline can be attributed to the information and engagement campaign and not to the other program components that were introduced later.

TABLE 2. Impact on Enrolment

	Enrolment	
Midline Enrolment		
Treatment	0.024** (0.012)	0.025** (0.012)
Controls		✓
Observations	820	820
Adjusted R-squared	0.004	0.051

Note: The table reports the marginal effect on enrolment. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.2. Mathematics. The results in Table 3 show girls in the treatment group experienced improvements in numeracy of 1.7 percentage points (0.06 SD) compared to the control group. Like the impact on enrolment, this statistically significant increase in mathematics performance occurred entirely between baseline and midline. This suggests that the information and engagement campaign had a significant, positive impact on the math performance of the girls.

To explore the gains in math performance in more detail, we consider the impact on the performance of girls in each of the EGMA subtasks. Girls' scores on the number quantities experienced the biggest increase, followed by addition, with 2.6 and 2.4 percentage point gains, respectively. We cannot distinguish whether the improvements in mathematics performance caused by the information and engagement campaign come from improvements in the understanding of mathematical concepts, or, alternatively,

TABLE 3. Impact on Mathematics Outcomes

	Number Identification EGMA 1	Number Quantities EGMA 2	Missing Numbers EGMA 3	Addition EGMA 4	Subtraction EGMA 5	Average Total
ML - BL						
Treatment	0.007 (0.011)	0.026** (0.0205)	0.013 (0.009)	0.024* (0.014)	0.010 (0.0123)	0.017* (0.009)
Controls	✓	✓	✓	✓	✓	✓
Observations	942	942	942	942	942	942
Adjusted R-Squared	0.312	0.347	0.342	0.412	0.409	0.532

Note: The table reports the coefficient on an indicator for belonging to an IGATE treatment school. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

from changes in attitudes or increases in confidence leading to an improved ability to apply their understanding under time pressure.²²

However, to further explore the mechanism through which girls improve their mathematics performance following the information and engagement campaign, the Appendix considers the impact of the IGATE program on the number of questions attempted on each subtest in the mathematics assessment. We do not observe any significant improvement in the number of questions students attempt, suggesting students have gained better mastery of the questions they attempt.²³

4.3. Literacy. When we examine the impact of IGATE on reading skills, we see no improvements between baseline and midline (see Table 4). This means that the information and engagement campaign promoting girls' education did not have the same short-run impacts on literacy that they did on numeracy performance.

4.4. Changes after Midline. By the time of the endline analysis a wide set of additional programs which directly provided resources to the communities had been implemented. In this section we consider the total impact of these additional interventions.

For enrollement we find no similar impact on enrolment between baseline and endline overall, suggesting that the subsequent interventions resulted in no additional improvement in enrolment.

In terms of mathematics, results in Table 6 indicate that the later intervention components did not lead to significant increases in scores on numeracy tests. We also find that the inclusion of midline test scores did not meaningfully change the baseline to endline

²²For example, Cotton et al. (2013) finds that gender gaps in mathematics performance depend at least partially on time constraints and competitive pressure. While EGMA 1, 4, and 5 do have a timed component, we do not find that the program significantly improved girls' performance under time pressure.

²³See Table A5 in the Appendix.

TABLE 4. Impact on Reading Outcomes

	Letter Sound Identification EGRA 1	Invented Words EGRA 2	Oral Fluency EGRA 3/4	Reading Comprehension EGRA 5	Average Total
ML - BL					
Treatment	-0.011 (0.012)	-0.003 (0.005)	0.004 (0.011)	-0.005 (0.021)	-0.004 (0.009)
Controls	✓	✓	✓	✓	✓
Observations	675	675	675	675	675
Adjusted R-Squared	0.126	0.370	0.553	0.306	0.506
Timed	Yes	Yes	Yes	No	
Early Stop Rule	Yes	Yes	Yes	No	

Note: The table reports the coefficient on an indicator for belonging to an IGATE treatment school. Note that EGRA 3 and 4 both assess oral fluency, but EGRA 3 was only given to girls who were in grades 1-5 at baseline while EGRA 4 was given to girls in who were in grades 6 and above at baseline. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 5. Impact on Enrolment

	Enrolment	
Endline Enrolment		
Treatment	0.012 (0.019)	0.012 (0.019)
Controls		✓
Observations	820	820
Adjusted R-squared	-0.001	0.073

Note: The table reports the marginal effect on enrolment. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

analysis. This indicates that the midline improvements in numeracy are not meaningfully mediating subsequent improvements in numeracy by endline (see Appendix C for details on the steps required to establish mediation).

However, in reading gains do occur overall between endline analysis with a significant improvement observed in letter sound identification of 2.4 percentage points, and improvement in reading comprehension of 6.4 percentage points. This contributed to an overall gain of 2.6 percentage points between baseline and endline. By endline, non-information-based interventions such as Happy Readers were in place. Among other things, this involved providing books and reading materials to schools. We know from other contexts that such resource provision, when targeted effectively, can improve student learning (Lavy, 2020; Snilstveit et al., 2016). It is therefore likely that this contributed to the reading improvements we observe at endline.

TABLE 6. Impact on Mathematics Outcomes

	Number Identification EGMA 1	Number Quantities EGMA 2	Missing Numbers EGMA 3	Addition EGMA 4	Subtraction EGMA 5	Average Total
ML - BL						
EL - BL						
Treatment	0.011 (0.010)	0.023 (0.015)	0.019 (0.015)	0.019 (0.014)	0.006 (0.017)	0.016 (0.011)
Controls	✓	✓	✓	✓	✓	✓
Midline test scores						
Observations	794	794	794	794	794	794
Adjusted R-Squared	0.172	0.172	0.260	0.340	0.335	0.401
EL - BL						
Treatment	0.005 (0.007)	0.007 (0.013)	0.010 (0.013)	0.004 (0.012)	-0.003 (0.013)	0.001 (0.007)
Controls	✓	✓	✓	✓	✓	✓
Midline test scores	✓	✓	✓	✓	✓	✓
Observations	794	794	794	794	794	794
Adjusted R-Squared	0.530	0.354	0.420	0.534	0.583	0.737

Note: The table reports the coefficient on an indicator for belonging to an IGATE treatment school. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

These improvements remain relatively unchanged in magnitude after adding in controls for midline test scores, again suggesting that the overall improvements in literacy are not meaningfully impacted by the program's impact on literacy skills.²⁴

However, it is also possible that the eventual gains in literacy occurred because the information-based interventions that were offered before midline took more time to be realized, or because the increased awareness developed by the initial wave of interventions made girls and community members more receptive to later interventions. This possibility is examined in more detail using a formal mediation analysis in Appendix C. However, it is also likely that they were at least in part driven by the other intervention components introduced after the midline data collection.

4.5. Attitudes. The theories of change that motivate the IGATE program and other interventions assume that dialogue drives changes in beliefs and norms, which in turn lead to positive changes in behavior. However, the GEC collected little data on attitudes and beliefs for IGATE project, presenting a limitation for the analysis. While we show that

²⁴This is not surprising, as the requirements for mediation have not been met (see Appendix C for a description of the steps required for mediation analysis). Since there is no impact on reading skills by midline, the second step of the four-step process for mediation analysis outlined by Judd & Kenny (1981), James & Brett (1984), Baron & Kenny (1986), and VanderWeele (2016), has not been satisfied. This step requires establishing that the independent variable (IGATE treatment) impacts the mediator (midline test scores). Table 4 shows that this condition for mediation is not satisfied. Similarly in the analysis of numeracy impacts, the first step of mediation analysis is not satisfied. More specifically, the treatment is not shown to have an impact on endline numeracy scores.

TABLE 7. Impact on Reading Outcomes

	Letter Sound Identification EGRA 1	Invented Words EGRA 2	Oral Fluency EGRA 3/4	Reading Comprehension EGRA 5	Average Total
EL - BL					
Treatment	0.030*** (0.012)	0.009 (0.005)	0.012 (0.007)	0.058*** (0.020)	0.028*** (0.007)
Controls	✓	✓	✓	✓	✓
Midline test scores	✓	✓	✓	✓	✓
Observations	598	598	598	598	598
Adjusted R-Squared	0.299	0.545	0.808	0.453	0.710
Timed	Yes	Yes	Yes	No	
Early Stop Rule	Yes	Yes	Yes	No	

Note: The table reports the coefficient on an indicator for belonging to an IGATE treatment school. Note that EGRA 3 and 4 both assess oral fluency, but EGRA 3 was only given to girls who were in grades 1-5 at baseline while EGRA 4 was given to girls in who were in grades 6 and above at baseline. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

dialogue leads to improvement in numeracy performance and academic progression, we do not have sufficient data on beliefs and attitudes to directly establish that dialogues improve beliefs, attitudes, or social norms of students, parents, teachers, or others.

The baseline attitude questions were limited to caregiver survey questions that asked whether the caregiver had positive aspirations for their girls, and whether they believed that girls can achieve as much as boys. Nearly all parents agreed with these statements, with 99% of caregivers reporting positive aspirations and 97% reporting beliefs that girls can achieve as much or more than their male peers, leaving no room to identify improvements.²⁵

The project added an additional measure of girls' attitudes and beliefs at later evaluation points with the Youth Leadership Index assessment being incorporated into girls' surveys. The midline and endline averages across the treatment and control groups are shown in Table A7 in the Appendix. However, this measure does not allow us to identify changes in these measures caused by the program itself. In addition, the sample size is not large enough to rule out that the differences at midline are not due random variation in the locations sampled between the two groups. Given these data limitations, we are unable to cleanly estimate the impact of IGATE on attitudes and beliefs directly.

To gain insight into the impact of the program on attitudes, we instead look for indicators that reveal increases in support for girls at the school or household levels. As we show in Table 8, when caregivers of girls who were around the age of menarche were asked whether they had purchased sanitary products for these girls in the past 12 months, caregivers in IGATE treatment areas were 3 percentage points more likely to have reported doing so than those in the control locations. This is a relevant observed

²⁵See Table A7 in the Appendix.

behavior, given the emphasis that access to sanitary wear received in the information and engagement campaign.

The change is not observable until after midline, with the midline control not meaningfully changing the endline results. However, since the interventions that were added after midline were limited to curriculum changes and resources that targeted girls, and not their caregivers or communities, it is likely that this overall improvement in caregiver attitudes can be largely attributed to the dialogue-based engagement campaign that included discussion around barriers girls face from menstruation and access to sanitary products. This finding suggests that the program has not only made households aware of the barriers girls face due to menstruation, but has also motivated them to take action to mitigate these barriers to support girls' education, and is consistent with other evidence on the efficacy of empowerment campaigns that focused on barriers related to menstrual health (Bandiera et al., 2020).

TABLE 8. Impact on Caregiver behavior

Caregiver purchased sanitary products for girl in past 12 months	
ML - BL	
Treatment	0.014 (0.035)
Controls	✓
Observations	595
Adjusted R-squared	0.180
EL - BL	
Treatment	0.038*** (0.014)
Midline behavior	
Controls	✓
Observations	545
Adjusted R-squared	0.886

Note: The table reports the coefficients for the linear probability model estimated by OLS. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Furthermore, Nordstrom (2021) presents additional evidence that dialogue-based intervention components affect attitudes and beliefs around girls' education in such environments using data from later waves of GEC programming in Zimbabwe. The GEC data collection around this later project included more detailed data with which to measure attitudes, allowing the analysis to show how a multifaceted gendered education

project that included dialogue-based engagement along with other types of intervention components significantly improved community attitudes towards girls' education. As with other GEC projects, however, the simultaneous implementation of the different components of the project Nordstrom (2021) studies does not allow one to isolate the impact of the dialogue campaigns from other program components, which our paper is able to do.

4.6. Further Considerations. There are several considerations to keep in mind when interpreting the results.

First, we are not able to determine the precise channel through which dialog-based engagement campaigns on girls' education lead to improvements in education outcomes. This is for two primary reasons. (I) Because the program simultaneously engaged groups of parents, girls, teachers, and community members, we cannot, for example, separately identify whether engaging girls or parents were more important for the overall effect of the program. (II) Because the GEC did not collect sufficient data on attitudes and beliefs, we cannot assess the degree to which the attitudes or support changed across different groups relative to others. Our subsection on attitudes provides some suggestive evidence that dialogue-based campaigns and other program components change attitudes and beliefs, but more work is needed.

Second, Zimbabwe is a country in which girls tend to perform fairly well compared to their peers in many other African countries and compared to boys in their own country.²⁶ This may suggest that there is less room for improvement in the academic performance of primary school girls in Zimbabwe than there may be in countries with substantial gender gaps. It could also imply that communities were already supportive of girls' education and providing information about the benefits of girls' education may not be a particularly effective policy. Both of these factors would make it less likely that providing information about the benefits of girls' education and the rights of girls will lead to substantial improvements in education outcomes. Yet, we still show this in the Zimbabwe context, suggesting that the impacts may be even larger in other countries.

Third, because the GEC only collected data on girls and not boys for the IGATE project, our analysis can only speak to the impact of the project on the absolute performance of girls and not on the impacts on boys or the overall gender gaps. We do not know, for

²⁶On the country's National Grade 7 exam, girls were more likely than their male peers to pass all four subject areas, with pass rates of 45% versus 40%, respectively Zimbabwe Schools Examination Council (2016). Primary school enrolment rates tend to be high for both girls and boys, approaching 99% according to The World Bank (2019), with girls demonstrating slightly higher completion rates than boys in the transition to secondary school. It is not until secondary school that gender gaps favoring males tend to develop. After the transition to secondary school, fewer females go on to graduate, with 19% of male adults completing upper secondary school versus only 12% of females (Chinembiri, 2018).

example, whether the academic performance of boys fell as parents and teachers put more emphasis on girls' education.

5. CONCLUSION

The current paper makes several contributions to the literature.

First, the analysis reports results from a randomized multifaceted education project implemented as part of the UK government's Girls' Education Challenge development portfolio. We show how the program led to improvements in academic progression, as well as literacy and mathematics performance among girls, despite being implemented in an environments where primary girls were already performing relatively well compared other settings.

Second, the analysis takes advantage of the staggered implementation of program components during implementation in order to identify the impact of the dialogue-based engagement components of the project. These components were the only portion of the project widely implemented ahead of the midline data collection, offering a unique opportunity to assess the impact of the dialogue components, which are typically implemented simultaneously with other types of programming during projects. We find that dialogue-based engagement was responsible for all of the observed improvements in mathematics performance and academic progression among participants, and that that they contributed to gains in literacy performance.

It may be easy to see how public discourse on girls' education could improve progression in school, which likely represent increases in attendance and engagement. Less obvious, perhaps, is how discourse may result in substantial improvements in mathematics performance over a relatively short time period. This may reflect additional effort by teachers and parents in helping girls learn the material, or additional effort by girls in studying or when taking the assessments. Future work may disentangle the mechanisms behind these improvements.

An important implication of these results concerns the motivation behind dialogue-based engagement campaigns. Often, dialogues are viewed as ways of changing social norms so that other program components will be more effective. Our results are consistent with this, but they additionally show how dialogue can improve outcomes even before other program components are implemented.

Third, these implications are not limited to education settings. Our paper presents some of the first robust quantitative evidence that dialogue-based engagement campaigns in any setting can have meaningful impacts on behavior and outcomes. Dialogues are increasingly incorporated as part of development and social programming. However, their inclusion is primarily supported by social change theories and qualitative

accounts rather than robust quantitative evidence. By establishing a causal link between public dialogue intervention components and targeted outcomes, our findings suggest that dialogue-based engagement can be effective at bringing about change, which may have applications for many settings. Further study is needed to better understand the effectiveness of such dialogues in and beyond education.

REFERENCES

- Acevedo, P., Cruces, G., Gertler, P., & Martinez, S. (2017). *Living up to expectations: How job training made women better off and men worse off* (Tech. Rep.). National Bureau of Economic Research.
- Acharya, A., Blackwell, M., & Sen, M. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review*, 110(3), 512–529.
- Alan, S., Ertac, S., & Mumcu, I. (2018). Gender stereotypes in the classroom and the effects on achievement. *Review of Economics and Statistics*, 100(5), 876-890.
- Andrabi, T., Das, J., & Khwaja, A. I. (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review*, 107(6), 1535-1563.
- Asadullah, M. N., & Chaudhury, N. (2009). Reverse gender gap in schooling in bangladesh: Insights from urban and rural households. *Journal of Development Studies*, 45(8), 1360-1380.
- Attanasio, O., & Kaufmann, K. (2014). Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender. *Journal of Development Economics*, 109, 203-216.
- Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., & Sulaiman, M. (2020). Women's empowerment in action: Evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics*, 12(1), 210–59.
- Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., & Khemani, S. (2010). Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy*, 2(1), 57-83.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173.
- Barrera-Osorio, F., Gonzalez, K., Lagos, F., & Deming, D. (2020). Providing performance information in education: An experimental evaluation in Colombia. *working paper*.
- Benbow, C., & Stanley, J. (1980). Sex differences in mathematical ability: fact or artifact? *Science*, 210(4475), 1262-1264.

- Berlinski, S., Busso, M., Dinkelman, T., & Martinez, C. (2016). Reducing parent-school information gaps and improving education outcomes: Evidence from high frequency text messaging in Chile. *working paper*.
- Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonsatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *Quarterly Journal of Economics*, 137(3), 1205-1242.
- Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355(6323), 389-391.
- Burgess, S., & Greaves, E. (2013). Test scores, subjective assessments, and stereotyping of ethnic minorities. *Journal of Labor Economics*, 31(3), 535-576.
- CARE. (2007). Ideas and action: Addressing the social factors that influence sexual and reproductive health.
(https://www.care.org/wp-content/uploads/2020/05/social_analysis_manual.pdf)
- CARE. (2014). Social Analysis and Action Methodology.
(https://www.care.org/wp-content/uploads/2020/08/final_saa_overview_brief.pdf)
- Chang, W., Diaz-Martin, L., Gopalan, A., Guarnieri, E., Jayachandran, S., & Walsh, C. (2020). What works to enhance women's agency: Cross-cutting lessons from experimental and quasi-experimental studies. *J-PAL Working Paper*.
- Chasin, R., Herzig, M., Roth, S., Chasin, L., Becker, C., & Robert R. Stains, J. (1996). From diatribe to dialogue on divisive public issues: Approaches drawn from family therapy. *Conflict Resolution*, 13(4), 323-344.
- Chinembiri, T. (2018). *The gender gap report* [Insight Report].
- Cislaghi, B., Denny, E. K., Cissé, M., Gueye, P., Shrestha, B., Shrestha, P. N., ... Clark, C. J. (2019). Changing social norms: the importance of "organized diffusion" for scaling up community health promotion and women empowerment interventions. *Prevention Science*, 20(6), 936-946.
- Cook, A. L., Shah, A., Brodsky, L., & Morizio, L. J. (2017). Strengthening school-family-community engagement through community dialogues. *Journal for Social Action in Counseling & Psychology*, 9(1), 9-37.
- Cortes, K. E., Fricke, H., Loeb, S., & Song, D. S. (2018). Too little or too much? Actionable advice in and early-childhood text messaging experiment. *NBER working paper* 24827.
- Cotton, C., McIntyre, F., & Price, J. (2013). Gender differences in repeated competition: Evidence from school math contests. *Journal of Economic Behavior & Organization*, 86, 52-66.
- De Mel, S., McKenzie, D., & Woodruff, C. (2011). Getting credit to high return microentrepreneurs: The results of an information intervention. *World Bank Economic Review*, 25(3), 456-485.

- Dessel, A., Rogge, M. E., & Garlington, S. B. (2006). Using intergroup dialogue to promote social justice and change. *Social Work, 51*(4), 303–315.
- Dhar, D., Jain, T., & Jayachandran, S. (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in india. *American Economic Review, 112*(3), 899–927.
- Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. *American Economic Review, 109*(8), 2728–2765.
- Doss, C., Fahle, E. M., Loeb, S., & York, B. N. (2018). More than just a nudge: Supporting kindergarten parents with differentiated and personalized text messages. *Journal of Human Resources, 54*(3), 567–603.
- Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature, 50*(4), 1051–79.
- Dupas, P. (2011). Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics, 3*(1), 1–34.
- Figueroa, M. E., Poppe, P., Carrasco, M., Pinho, M. D., Massingue, F., Tanque, M., & Kwizera, A. (2016). Effectiveness of Community Dialogue in Changing Gender and Sexual Norms for HIV Prevention: Evaluation of the Tchova Tchova Program in Mozambique. *Journal of Health Communications, 21*(5), 554–562.
- Fitzsimons, E., Malde, B., Mesnard, A., & Vera-Hernandez, M. (2016). Nutrition, information and household behavior: Experimental evidence from Malawi. *Journal of Development Economics, 122*, 113–126.
- Friedman, T., Schwantner, U., Spink, J., Tabata, N., & Waters, C. (2016). Improving quality education and children's learning outcomes. *UNICEF report*.
- Fryer, R. G., Jr., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics, 2*(2), 210–240.
- GEC. (2018). Thematic Review: Community Based Awareness Attitudes and Behavior. *UKaid Girls' Education Challenge*.
- Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The homecoming of American college women: The reversal of the college gender gap. *Journal of Economic Perspectives, 20*(4), 133–156.
- Gudyanga, A. (2016). Zimbabwean female participation in physics: Factors of identity formation considered as contributing to developing an orientation to physics by female students. *Journal of Education and Practice, 7*(26), 159–171.
- Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Culture, gender, and math. *Science, 320*(5880), 1164–1165.
- Hastings, J. S., & Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics, 123*(4),

1373-1414.

- Hedges, L., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, 269(5220), 41-45.
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008). Gender similarities characterize math performance. *Science*, 321(5888), 494-495.
- James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of applied psychology*, 69(2), 307.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2), 515-548.
- John, A., & Orkin, K. (2022). Can simple psychological interventions increase preventive health investment? *Journal of the European Economic Association*, 20(3), 1001-1047.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation review*, 5(5), 602-619.
- Krämer, M., Kumar, S., & Vollmer, S. (2019). Anemia, diet, and cognitive development: Impact of health information on diet quality and child nutrition in rural India. *Courant Research Centre: Poverty, equity and growth discussion paper*.
- Lavecchia, A. M., Oreopoulos, P., & Brown, R. S. (in press). Long-run effects from comprehensive student support: Evidence from Pathways to Education. *American Economic Review: Insights*.
- Lavy, V. (2020). Expanding school resources and increasing time on task: Effects on students' academic and noncognitive outcomes. *Journal of the European Economic Association*, 18(1), 232-265.
- Lavy, V., & Sand, E. (2018). On the origins of the gender human capital gap: Short- and long-term consequences of teachers' biases. *Journal of Public Economics*, 167, 263-279.
- Martin, S., ao, J. L., Muhangi, D., Nuwa, A., Magul, D., & Counihan, H. (2017). Community dialogues for child health: Results from a qualitative process evaluation in three countries. *Journal of Health, Population and Nutrition*, 36. (article number 29)
- Ndagije, H. B., Manirakiza, L., Kajungu, D., Galiwango, E., Kusemererwa, D., Olsson, S., ... Speybroeck, N. (2019). The effect of community dialogues and sensitization on patient reporting of adverse events in rural Uganda: Uncontrolled before-after study. *PLOS One*, 14(5). (e0203721)
- Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from Madagascar. *JPAL Working Paper*.
- Nollenberger, N., Rodriguez-Planas, N., & Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5), 257-261.
- Nordstrom, A. (2021). It takes a village: The impact of community mobilization campaigns on attitudes and education. *Queen's University Economics Working Papers*.

- Nordstrom, A., & Cotton, C. (2020). Impact of a severe drought on education: More schooling does not imply more learning. *Queen's University Working Paper*.
- OECD. (2015). *The ABC of Gender Equality in Education* (Tech. Rep.).
- Oreopoulos, P., Brown, R. S., & Lavecchia, A. M. (2017). Pathways to Education: An integrated approach to helping at-risk high school students. *Journal of Political Economy*, 125(4), 947-984.
- Robinson-Cimpian, J. P., Lubinski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Development Psychology*, 50(4), 1262-1281.
- Rodriguez-Planas, N., & Nollenberger, N. (2018). Let the girls learn! it is not only about math ... it's about gender social norms. *Economics of Education Review*, 62, 230-253.
- Rosenzweig, M. R., & Zhang, J. (2013). Economic growth, comparative advantage, and gender differences in schooling outcomes: Evidence from the birthweight differences in Chinese twins. *Journal of Development Economics*, 104, 245-260.
- Schweizer, S., Davis, S., & Thompson, J. L. (2013). Changing the conversation about climate change: A theoretical framework for place-based climate change engagement. *Environmental Communication*, 7(1), 42-62.
- Shrestha, S. A., & Yang, D. (2019). Facilitating worker mobility: A randomized information intervention among migrant workers in Singapore. *Economic Development and Cultural Change*, 68(1), 63-91.
- Singh, A. (2020). Learning more with every year: School year productivity and international learning divergence. *Journal of the European Economic Association*, 18(4), 1770-1813.
- Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., ... Jimenez, E. (2016). The impact of education programmes on learning and school participation in low-and middle-income countries.
- The World Bank. (2019). *Zimbabwe*. (data retrieved from the World Bank, <https://data.worldbank.org/country/zimbabwe>)
- Unterhalter, E., North, A., Arnot, M., Lloyd, C., Moletsane, L., Murphy-Graham, E., ... Saito, M. (2014). Education Rigorous Literature Review: Interventions to Enhance Girls' Education and Gender Equality. *UKaid/DFID commissioned report*.
- VanderWeele, T. J. (2016). Mediation analysis: a practitioner's guide. *Annual review of public health*, 37, 17-32.
- Walsh, M. E., Madaus, G. F., Raczek, A. E., Dearing, E., Foley, C., An, C., ... Beaton, A. (2014). A new model for student support in high-poverty urban elementary schools: Effects on elementary and middle school academic outcomes. *American Educational Research Journal*, 51(4), 704-737.

- Wegs, C., Creanga, A. A., Galavotti, C., & Wamalwa, E. (2019). Community Dialogue to Shift Social Norms and Enable Family Planning: An Evaluation of the Family Planning Results Initiative in Kenya. *PLOS One*, 11(4). (e0153907)
- Zimbabwe Schools Examination Council. (2016). *November 2015 Grade 7 Results Analysis Report* (Tech. Rep.). Retrieved from <http://www.zimsec.co.zw/november-2016-grade-7-results-analysis-report/>

APPENDIX A. APPENDIX

A.1. School Locations. Figure A1 shows a map of the IGATE school locations across rural districts in Zimbabwe.

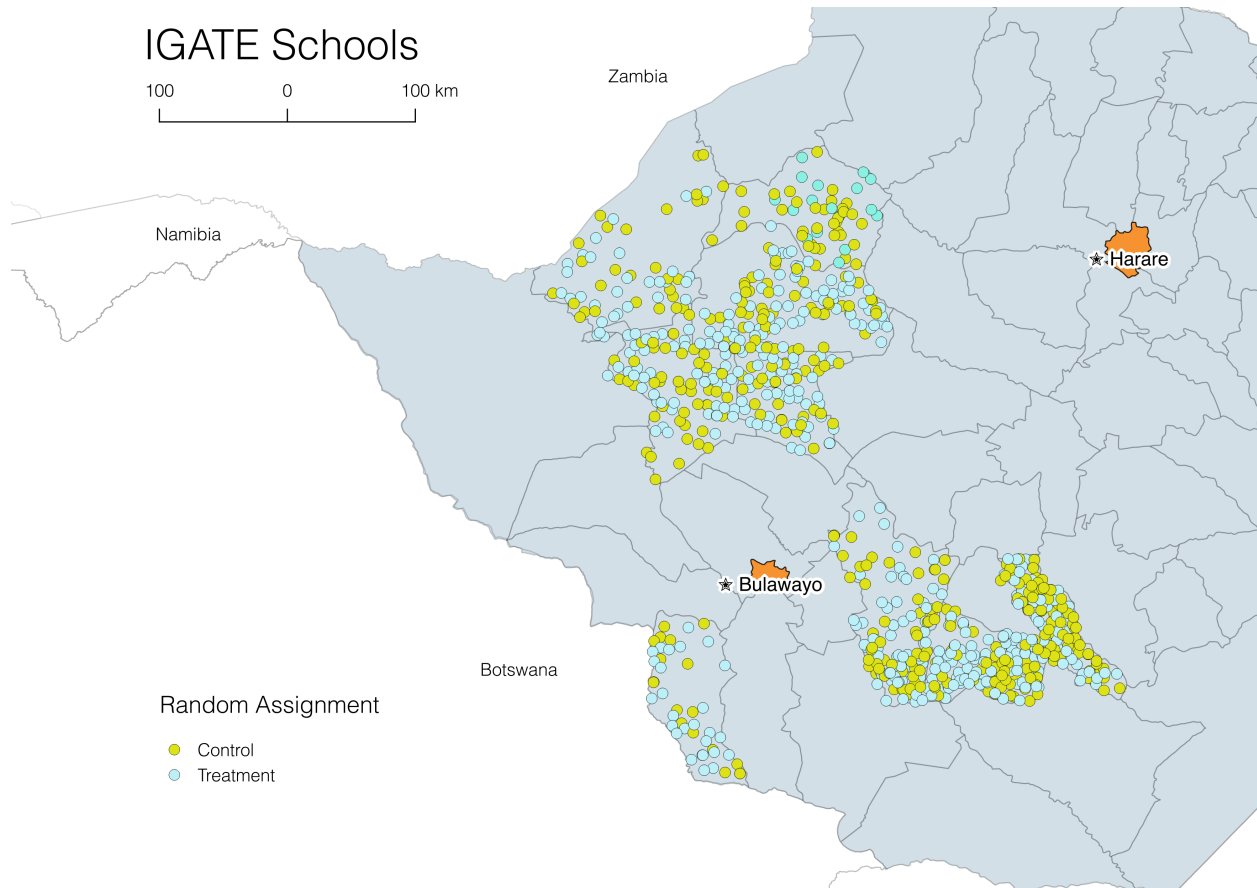


FIGURE A1. IGATE School Locations

APPENDIX B. EGRA/EGMA TEST DETAILS

As is standard with the EGRA and EGMA assessments, the questions in each test were described verbally, one by one, by a professional enumerator to individual students. Students then provided their answers verbally and enumerators record whether the student's answer was correct. During the test, students are given visual stimuli to follow along and to see the specific numbers, letters, and words they are asked to say or analyze. There are five subtasks that make up the numeracy assessment: number identification, number quantities, missing numbers, addition, and subtraction. The number identification subtask consists of 20 numbers which students are asked to identify in one minute. An example of a typical EGMA number identification subtask as viewed by the enumerator is shown in figure A2.

Start					54
9	5	0	26	52	
69	92	61	53	86	*
57	44	32	28	17	*
686	753	914	829	234	*
Stop					

FIGURE A2. EGMA 1 (Number Identification) Example

In the quantity discrimination subtask, a student is presented with a list of pairs of numbers and is asked to identify the larger number. This exercise is not timed but ends after 4 incorrect answers in a row or a hesitation of 5 seconds by the student. This stop rule trigger is shown in figure A3.

Correct Incorrect No response

3) What number goes here?
_, 20 , 30, 40
[10]

Correct Incorrect No response

4) What number goes here?
200, 300, 400, _
[500]

Correct Incorrect No response

5) What number goes here?
8, 10, _, 14
[12]

Correct Incorrect No response

6) What number goes here?

Auto skipped

FIGURE A3. EGMA 3 (Missing Numbers) Example: Early Stop Rule Trigger

The addition and subtraction level 1 components include 20 problems each. According to the EGMA guidelines, subtraction questions must be the inverse of the additional questions. A stop rule after 5 incorrect answers applies to these tasks as well.

There are five literacy subtasks: letter and sound identification, invented words, oral fluency (grades 1-5, and grades 6-7), and reading comprehension. The letter and sound identification tasks involve students phonetically reading individual letters in the alphabet, much like the number identification task. The enumerator records each correct pronunciation. Invented word tasks involve 50 words that do not have a meaning in English or in the local languages. The student is asked to read each made-up word aloud and the enumerator records each correct pronunciation. An example of an invented words subtask as viewed by the enumerator is shown in figure A4.

Start		
dur	zid	Hib
boq	kaz	Cog
rit	nak	jol
jev	Yot	muk

FIGURE A4. EGRA 2 (Invented Words) Example

The remaining subtasks, oral fluency and reading comprehension, ask students to read a short story aloud. Enumerators are instructed to record the words the students misidentified or mispronounced and to identify the last word the student correctly said aloud within the time limit. The reading comprehension task then asks the students questions about the passage to assess their understanding of the story they just read.

The test design guidelines specify all details about each question's difficulty level. This includes details about the number each sequence increases by in numeracy subtasks and the number of single, double, and triple-digit numbers to be used in the Missing Numbers and Number Identification subtasks; subtraction problems are required to be the inverse of the addition problems. In the first two EGRA components, the versions are differentiated by reordering letters or words within the rows to retain the same level of difficulty. The EGRA story subtasks are written with the intention of remaining the same difficulty using the same number of words per sentence and per passage and using a similar vocabulary. Given this strict structure, different versions of the tests are not likely to be different difficulties.

B.1. Additional Treatment and Control Comparison Tables. Here, we provide additional tables comparing the treatment and control group baseline characteristics, illustrating that there are no substantial differences between the groups ahead of the IGATE implementation.

TABLE A1. Numeracy Test Subtasks - Attrited versus Remaining Samples

	Attrited		Remaining	
	Control	Treatment	Control	Treatment
EGMA 1: Number Identification	0.745 (0.322)	0.705 (0.334)	0.693 (0.327)	0.677 (0.338)
% of questions attempted	0.933 (0.170)	0.956 (0.126)	0.951 (0.124)	0.955 (0.122)
% of time spent	0.694 (0.237)	0.748 (0.229)	0.733 (0.236)	0.737 (0.236)
EGMA 2: Number Quantities	0.656 (0.343)	0.604 (0.346)	0.571 (0.331)	0.557 (0.354)
% of questions attempted	0.928 (0.171)	0.899 (0.208)	0.902 (0.195)	0.876 (0.224)
EGMA 3: Missing Numbers	0.444 (0.281)	0.384 (0.246)	0.376 (0.260)	0.372 (0.266)
% of questions attempted	0.880 (0.205)	0.859 (0.201)	0.832 (0.222)	0.820 (0.232)
EGMA 4: Addition	0.485 (0.269)	0.424 (0.260)	0.427 (0.294)	0.409 (0.284)
% of questions attempted	0.712 (0.141)	0.670 (0.148)	0.710 (0.152)	0.688 (0.161)
% of time spent	0.991 (0.0504)	0.994 (0.0357)	0.986 (0.0630)	0.984 (0.0776)
EGMA 5: Subtraction	0.380 (0.256)	0.289 (0.231)	0.316 (0.273)	0.298 (0.250)
% of questions attempted	0.684 (0.149)	0.640 (0.158)	0.668 (0.146)	0.649 (0.161)
% of time spent	0.987 (0.0704)	0.992 (0.0512)	0.986 (0.0683)	0.990 (0.0629)
Average Numeracy Score	0.542 (0.261)	0.481 (0.247)	0.477 (0.265)	0.462 (0.266)
% of questions attempted	0.827 (0.0945)	0.805 (0.100)	0.813 (0.111)	0.798 (0.114)

TABLE A2. Baseline Summary Statistics - Attrited and Remaining Samples

	Attrited		Remaining	
	Control	Treatment	Control	Treatment
Age	9.375 (2.268)	9.574 (1.875)	9.377 (2.011)	9.380 (1.997)
Grade	3.781 (1.827)	3.607 (1.584)	3.676 (1.746)	3.627 (1.779)
Illness	0.0938 (0.296)	0.115 (0.321)	0.103 (0.305)	0.112 (0.316)
Disability	0.219 (0.420)	0.230 (0.424)	0.174 (0.380)	0.191 (0.394)
Orphan	0.0938 (0.296)	0.0164 (0.128)	0.0676 (0.252)	0.0536 (0.226)
Travel time to school (minutes)	27.56 (15.60)	33.20 (27.02)	32.96 (23.16)	35.12 (27.30)
Household often goes hungry	0.406 (0.499)	0.426 (0.499)	0.224 (0.418)	0.228 (0.420)
Household often goes thirsty	0.188 (0.397)	0.148 (0.358)	0.128 (0.335)	0.121 (0.327)
Caregiver has no education	0.0625 (0.246)	0.115 (0.321)	0.0676 (0.252)	0.0979 (0.298)
Caregiver has primary education	0.375 (0.492)	0.492 (0.504)	0.523 (0.500)	0.550 (0.498)
Caregiver has secondary education	0.562 (0.504)	0.393 (0.493)	0.409 (0.493)	0.352 (0.478)
Caregiver works outside of household	0.281 (0.457)	0.180 (0.388)	0.221 (0.415)	0.235 (0.425)

TABLE A3. Numeracy Test Subtasks - Summary Statistics

	Control	Treatment	Difference
EGMA 1: Number Identification	0.693 (0.327)	0.677 (0.338)	-0.016
% of questions attempted	0.951 (0.124)	0.955 (0.122)	0.004
% of time spent	0.733 (0.236)	0.737 (0.236)	0.004
EGMA 2: Number Quantities	0.571 (0.331)	0.557 (0.354)	-0.014
% of questions attempted	0.902 (0.195)	0.876 (0.224)	-0.026
EGMA 3: Missing Numbers	0.376 (0.260)	0.372 (0.266)	-0.004
% of questions attempted	0.832 (0.222)	0.820 (0.232)	-0.012
EGMA 4: Addition	0.427 (0.294)	0.409 (0.284)	-0.018
% of questions attempted	0.710 (0.152)	0.688 (0.161)	-0.022
% of time spent	0.986 (0.0630)	0.984 (0.0776)	-0.002
EGMA 5: Subtraction	0.316 (0.273)	0.298 (0.250)	-0.018
% of questions attempted	0.668 (0.146)	0.649 (0.161)	-0.019
% of time spent	0.986 (0.0683)	0.990 (0.0629)	0.004
Average Numeracy Score	0.477 (0.265)	0.462 (0.266)	-0.015
% of questions attempted	0.813 (0.111)	0.798 (0.114)	-0.015
N	385	557	

Note these numbers measure baseline levels for girls who could be recontacted at midline.

B.2. Questions Attempted on EGMA. Table A5 reports results from an analysis considering the impact of IGATE on the number of questions attempted on the EGMA exam. The first panel shows that the relevant coefficients from the DiD analysis of question attempts are positive and significant. This suggests that the community information interventions led girls to answer more questions between baseline and midline. Note that

TABLE A4. Sample Grade Distribution

Baseline Grade	Control	Treatment	Difference
1	15%	15%	0%
2	17%	17%	-1%
3	14%	18%	3%
4	19%	16%	-3%
5	21%	20%	-1%
6	8%	9%	0%
7	5%	6%	0%
N	385	557	

in the case of the number quantities subtask, this should be interpreted as an indication of improved ability rather than increased effort since this subtask was stopped after participants incorrectly answered four questions in a row.

TABLE A5. Percentage of Questions Attempted

	EGMA 1	EGMA 2	EGMA 3	EGMA 4	EGMA 5	Total
ML-BL						
Treatment	0.003 (0.003)	0.010 (0.008)	0.007 (0.010)	0.008 (0.012)	-0.001 (0.014)	0.008 (0.007)
Controls	✓	✓	✓	✓	✓	✓
Observations	942	942	942	942	942	942
Adjusted R-Squared	0.040	0.192	0.238	0.324	0.241	0.377
EL - BL						
Treatment	0.001 (0.002)	0.003 (0.006)	0.018** (0.009)	0.018** (0.009)	-0.0004 (0.016)	0.010 (0.007)
Controls	✓	✓	✓	✓	✓	✓
Midline attempt percent						
Observations	794	794	794	794	794	794
Adjusted R-Squared	0.020	0.098	0.158	0.158	0.182	0.288
EL - BL						
Treatment	0.001 (0.002)	-0.0002 (0.005)	0.015* (0.008)	0.003 (0.011)	-0.003 (0.014)	0.003 (0.005)
Controls	✓	✓	✓	✓	✓	✓
Midline attempt percent	✓	✓	✓	✓	✓	✓
Observations	794	794	794	794	794	794
Adjusted R-Squared	0.033	0.384	0.353	0.379	0.342	0.595
Timed	Yes	No	No	Yes	Yes	
Early Stop Rule	No	Yes	Yes	No	No	

Note: The table reports the coefficient on an indicator for belonging to an IGATE treatment school. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

APPENDIX C. MEDIATION ANALYSIS OF SEQUENTIAL LEARNING GAINS

The main results did not provide convincing evidence that the program's impact on midline literacy or numeracy mediated the respective endline results. However, this focus on within-domain improvements may overlook the broader association between earlier improvements in one domain (ie. literacy or numeracy) and later improvements in the other.

To measure the associations between earlier midline gains and later improvements in literacy at endline, Table A6 shows the results from a variation of equation 2, where the lagged dependent variables also include the student's overall midline and baseline numeracy test score. Following the mediation process outlined by Judd & Kenny (1981), James & Brett (1984), Baron & Kenny (1986), and VanderWeele (2016), a mediation analysis allows us to determine whether the student's gains in numeracy by midline either partially or completely mediate the impact we observe on literacy at endline. This involves a four-stage process. The first stage requires establishing that there is a causal relationship between IGATE and the ultimate outcome, either literacy or numeracy. The second stage requires establishing that there is a causal relationship between IGATE and the mediator, the student's skills. The third and fourth steps are done simultaneously by determining whether the independent variable's (ie. the IGATE treatment) impact on ultimate outcomes is either completely or partially mediated by the program. This is done analyzing the coefficient on a regression that includes both the independent variable and the mediator. If the midline test scores *completely* mediate the program's impact on endline test scores, then we would expect the coefficient on the IGATE indicator to be zero in this final specification.

Table A6 presents the results from steps 2, 3, and 4 of this process (step one has been established in the main results in the analysis of numeracy test scores at midline). We find that the endline impact observed on letter sound identification is completely mediated by the impact the program had already had at midline. This is evidenced by the fact that the coefficient on the treatment is no longer significantly different from 0 after adding in the mediator: the change in numeracy tests by midline. At the same time, the improvements in reading comprehension and the impact on reading overall are at least *partially* mediated by the impact the program had had at midline. This is evidenced by the fact that the coefficient on the treatment indicator is still significant after adding the mediator, while the midline test score coefficient is as well.

TABLE A6. Impact on Reading Outcomes - Mediation Analysis

	Letter Sound Identification EGRA 1	Invented Words EGRA 2	Oral Fluency EGRA 3/4	Reading Comprehension EGRA 5	Average Total
EL - BL (Mediation Step 2)					
Midline numeracy test scores	0.273*** (0.050)	0.291*** (0.025)	0.686*** (0.050)	0.943*** (0.093)	0.542** (0.065)
Midline literacy test scores					
Baseline literacy and numeracy test scores	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
Observations	597	597	597	597	597
Adjusted R-Squared	0.229	0.457	0.623	0.408	0.586
EL - BL (Mediation Steps 3 and 4)					
Treatment	0.018 (0.012)	0.004 (0.006)	0.012 (0.010)	0.041** (0.020)	0.019*** (0.007)
Midline numeracy test scores	0.269*** (0.039)	0.290*** (0.025)	0.683*** (0.050)	0.934*** (0.092)	0.019** (0.009)
Midline literacy test scores					
Baseline literacy and numeracy test scores	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
Observations	597	597	597	597	597
Adjusted R-Squared	0.230	0.456	0.623	0.453	0.588

Note: The table reports the coefficient on an indicator for belonging to an IGATE treatment school. Note that EGRA 3 and 4 both assess oral fluency, but EGRA 3 was only given to girls who were in grades 1-5 at baseline while EGRA 4 was given to girls in who were in grades 6 and above at baseline. Controls include baseline age, grade, and geographic district. Cluster-robust standard errors are in parentheses. Standard errors are clustered at the school level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

APPENDIX D. ATTITUDE MEASURES

TABLE A7. Caregiver and learner attitudes at baseline, midline, and endline

Baseline	Control	Treatment	Difference
Caregiver has positive aspirations for their child	0.979 (0.143)	0.991 (0.0945)	0.012
Caregiver believes girls can learn as much as boys	0.979 (0.143)	0.964 (0.186)	-0.015
Caregiver listens to daughter's views when making decisions about girl's education	0.662 (0.474)	0.646 (0.479)	-0.016
Midline			
Caregiver has positive aspirations for their child	0.966 (0.181)	0.995 (0.0684)	0.029
Caregiver believes girls can learn as much as boys	0.973 (0.162)	0.953 (0.212)	-0.02
Caregiver listens to daughter's views when making decisions about girl's education	0.591 (0.493)	0.724 (0.448)	0.133
Girl's youth leadership index score (/100)	0.669 (0.146)	0.673 (0.147)	0.004
Endline			
Caregiver has positive aspirations for their child	0.988 (0.110)	0.998 (0.0462)	0.01
Caregiver believes girls can learn as much as boys	0.976 (0.154)	0.968 (0.176)	-0.008
Caregiver listens to daughter's views when making decisions about girl's education	0.683 (0.466)	0.642 (0.480)	-0.041
Girl's youth leadership index score (/100)	0.656 (0.158)	0.670 (0.147)	0.014